Exhibit 14
Part 1

Physics of Semiconductor Devices

SECOND EDITION

S. M. Sze

Bell Laboratories, Incorporated
Murray Hill, New Jersey

A WILEY-INTERSCIENCE PUBLICATION

JOHN WILEY & SONS

New York • Chichester • Brisbane • Toronto • Singapore

f Congress Cataloging in Publication Data:

., 1936-
s of semiconductor devices.

ey-Interscience publication."
s index.
conductors. I. Title.

5.S988 1981 537.6'22 81-213
71-05661-8 AACR2

the United States of America

2

p-n Junction Diode

- INTRODUCTION
- BASIC DEVICE TECHNOLOGY
- DEPLETION REGION AND DEPLETION CAPACITANCE
- CURRENT–VOLTAGE CHARACTERISTICS
- JUNCTION BREAKDOWN
- TRANSIENT BEHAVIOR AND NOISE
- TERMINAL FUNCTIONS
- HETEROJUNCTION

2.1 INTRODUCTION

The *p-n* junctions are of great importance both in modern electronic applications and in understanding other semiconductor devices. The *p-n* junction theory serves as the foundation of the physics of semiconductor devices. The basic theory of current–voltage characteristics of *p-n* junctions was established by Shockley.[1] This theory was then extended by Sah, Noyce, and Shockley[2], and by Moll.[3]

In this chapter we first briefly discuss the basic device technology which is pertinent not only to *p-n* junctions but also to most semiconductor devices. Then the basic equations presented in Chapter 1 are used to develop the ideal static and dynamic characteristics of *p-n* junctions. Departures from the ideal characteristics due to generation and recombination in the depletion layer, to high injection, and to series resistance effects are then discussed. Junction breakdown (especially that due to avalanche multiplication) is considered in detail, after which transient behaviors and noise performance in *p-n* junction are presented.

A p-n junction is a two-terminal device. Depending on doping profile, device geometry, and biasing condition, a p-n junction can perform various terminal functions, which are considered briefly in Section 2.7. The chapter closes with attention to an important group of devices, the heterojunctions, which are junctions formed between dissimilar semiconductors (e.g., n-type Ge on p-type GaAs).

2.2 BASIC DEVICE TECHNOLOGY

In this chapter we are concerned primarily with silicon technology,[4] because its development is more advanced than that of any other semiconductor. Some important device fabrication methods are shown in Fig. 1.

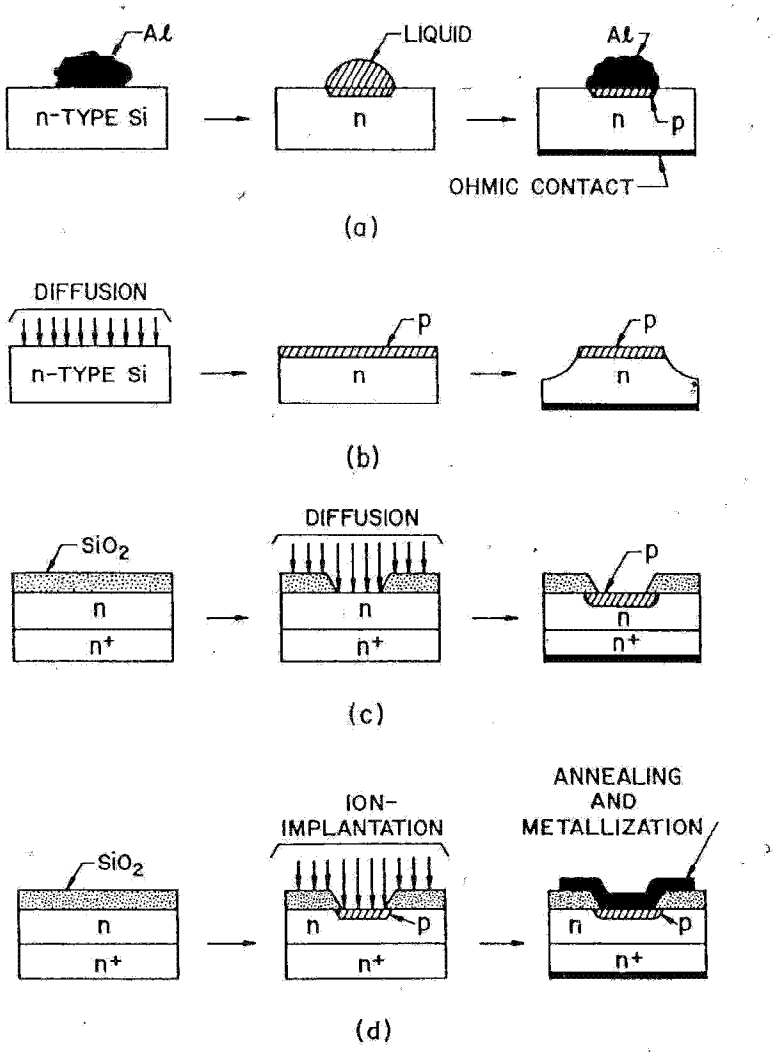

Fig. 1 Some device fabrication methods. (*a*) Alloyed junction. (*b*) Diffused mesa junction. (*c*) Diffused planar junction on epitaxial substrate. (*d*) Ion implantation.

In the alloy method,[5] Fig. 1*a*, a small pellet of aluminum is placed on an *n*-type ⟨111⟩-oriented silicon wafer. The system is then heated to a temperature slightly higher than the eutectic temperature (~580°C for the Al–Si system) so that a small puddle of molten Al–Si mixture is formed. The temperature is then lowered and the puddle begins to solidify. A recrystallized portion, which is saturated with the acceptor impurities and with the same crystal orientation, forms the heavily doped *p*-type region (p^+) on the *n*-type substrate. The aluminum button on top can be used as an ohmic contact for the *p*-type region. For the ohmic contact on the *n*-type wafer, a Au–Sb alloy (with ~0.1% Sb) can be evaporated onto the wafer and alloyed at about 400°C to form a heavily doped *n*-type region (n^+). On a *p*-type wafer, the roles of the aluminum and the Au–Sb alloy can be interchanged to form an n^+ junction on top and p^+ ohmic contact on the bottom of the wafer. The junction location obtained by the alloy method depends critically on the temperature–time alloying cycle and is difficult to control precisely.

The solid-state diffusion method was developed in 1956 to give more precise control of the impurity profile.[6] A diffused mesa junction method, where *p*-type impurities (e.g., boron in the form of BBr_3) are diffused into the *n*-type substrate, is shown in Fig. 1*b*. After diffusion, portions of the surface are protected (e.g., by wax or metal contacts), and the rest are etched out to form the mesa structures.

A new degree of control over the lateral geometry of the diffused junction is achieved by using an insulating layer that can prevent most donor and acceptor impurities from diffusing through it.[7] A typical example is shown in Fig. 1*c*. A thin layer of silicon dioxide (~1 μm) is thermally grown on silicon. With the help of lithographic techniques (e.g., photolithography, x-ray, or electron-beam lithography), portions of the oxide can be removed and windows (or patterns) cut in the oxide. The impurities will diffuse only through the exposed silicon surface, and *p*-*n* junctions will form in the oxide windows. This process, the planar process,[8] has become the principal method of fabricating semiconductor devices and integrated circuits.*

Also shown in Fig. 1*c* is the epitaxial substrate, for example, an *n* layer on n^+ substrate.[9] Such a substrate is generally used in the planar process to reduce series resistance. Epitaxy, derived from the Greek words *epi*, meaning "on", and *taxis*, meaning "arrangement", describes a technique of crystal growth by chemical reaction used to form, on the surface of crystal, thin layers of semiconductor materials with lattice structures identical to

*The planar process uses techniques which were previously known, such as oxide masking against diffusion. The distinguishing feature is their use in combination, permitting an unprecedented fineness in controlling the sizes and shapes of electrodes and diffused regions. The name "planar" comes from the requirement that the wafer surface must be approximately flat. If the surface is rough, the liquid photoresist does not coat it evenly, and imperfections result.

those of the crystal. In this method, lightly doped high-resistivity epitaxial layers are grown on and supported by a heavily doped low-resistivity substrate thus ensuring both the desired electrical properties and mechanical strength.

Figure 1*d* shows a *p-n* junction formed by ion implantation.[10] To date, this method gives the most precise control of an impurity profile. Ion implantation can be done at room temperature, and the implantation-induced lattice damages can be removed by annealing at about 700°C or less. Therefore, ion implantation is a relatively low-temperature process compared to diffusion, which is generally done at 1000°C or higher.

We shall now briefly discuss the four main processes of planar technology: epitaxial growth, oxidation, diffusion of impurities, and ion implantation.

Epitaxial layers can be formed by the vapor-phase growth technique.[4] The basic reaction that results in the growth of silicon layers is $SiCl_4 + 2H_2 \rightarrow Si(solid) + 4HCl(gas)$. Typically, the silicon layer is grown at a rate of about 1 μm/min at 1200°C or higher temperature with a mole fraction of $SiCl_4$ at 0.01% (ratio of $SiCl_4$ molecules to the total number of molecules in the gas). Epitaxial layers can also be formed by liquid-phase growth,[11] which has been extensively used for compound semiconductors, or by molecular-beam epitaxy,[12,13] which can give precise control of semiconductor compositions down to atomic dimensions.

The most frequently used method to form silicon dioxide films is by thermal oxidation[14] of silicon through the chemical reaction: $Si(solid) + O_2(dry\ oxygen) \rightarrow SiO_2(solid)$ or $Si(solid) + 2H_2O(steam) \rightarrow SiO_2(solid) + 2H_2$. It can be shown that for short reaction times the oxide thickness increases linearly with time, and for prolonged oxidation the thickness varies as the square root of time—the so-called parabolic relationship.[15] When a silicon dioxide film of thickness d is formed, a layer of silicon of thickness $0.45\,d$ is consumed. Figure 2 shows the experimental results of the oxide thickness as a function of reaction time and temperature for both dry oxygen growth and steam growth.

At a given temperature, the oxidation rate in steam is about 5 to 10 times higher than for dry oxygen. Also, at lower temperatures, the oxidation rates show pronounced dependence on crystal orientation.[16]

The simple one-dimensional diffusion process can be given by the Fick equation,[17]

$$\frac{\partial C(x,t)}{\partial t} = D\frac{\partial^2 C(x,t)}{\partial x^2} \tag{1}$$

where C is the impurity concentration and D the diffusion coefficient. This expression is similar to that given in Eq. 96a of Chapter 1, without generation, recombination, or electric field. For a "limited source" condition, with the total amount of impurities S, the solution of Eq. 1 is given

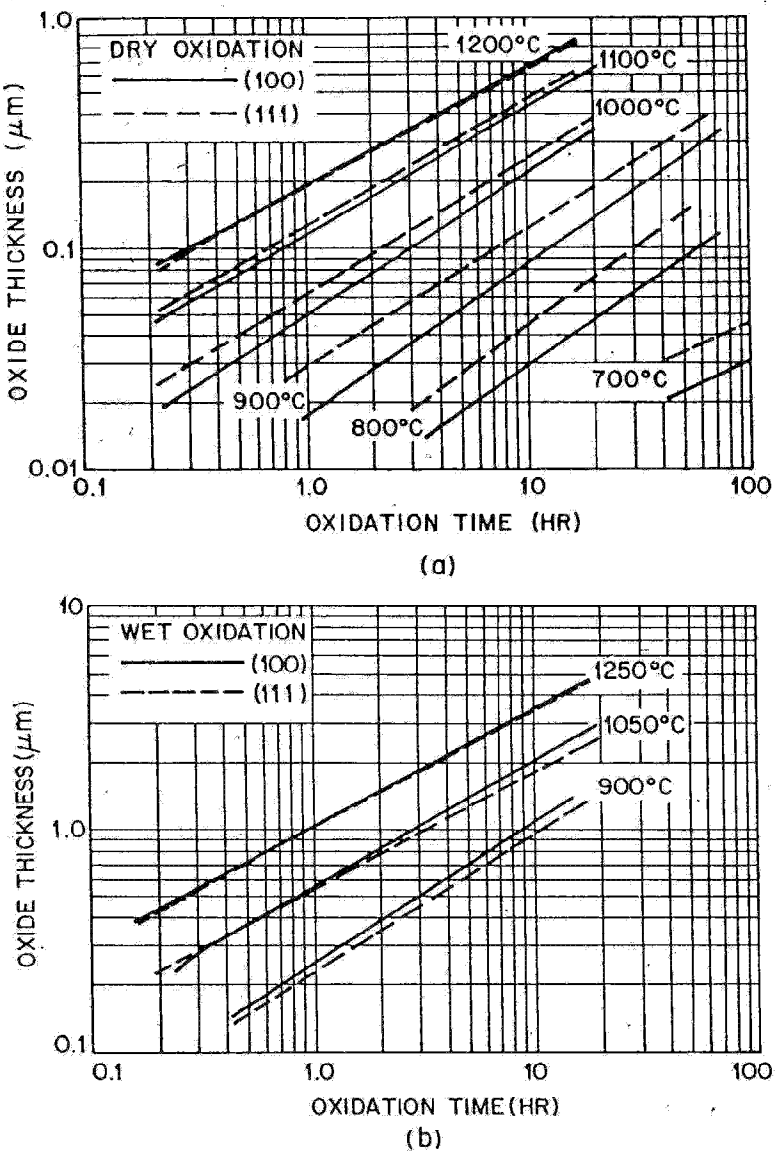

Fig. 2 Experimental results of silicon dioxide thickness as a function of the reaction time and temperature for two substrate orientations. (*a*) Dry oxygen growth. (*b*) Steam growth. (After Meindl et al., Ref. 16.)

by the Gaussian function

$$C(x,t) = \frac{S}{\sqrt{\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right). \tag{2a}$$

For the "constant surface concentration" condition with a surface concentration C_s, the solution of Eq. 1 is given by the error function complement

$$C(x,t) = C_s \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right). \tag{2b}$$

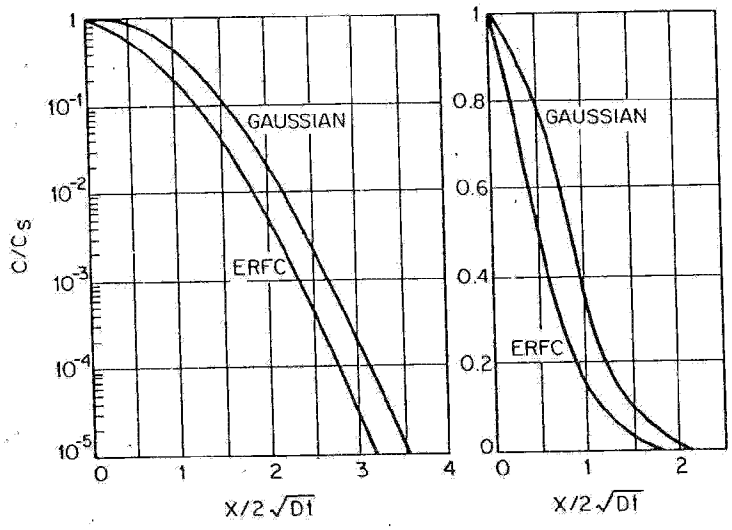

Fig. 3 Normalized concentration versus normalized distance for Gaussian and error function complement (erfc) distributions plotted in both semilog and linear scales. (After Carslaw and Jaeger, Ref. 17.)

The normalized concentration versus normalized distance to the foregoing two solutions is shown in Fig. 3. The diffusion profiles of many impurities can indeed be approximated by the preceding expressions. Many, however, have more complicated profiles, for example, As in Si with a diffusion process depending strongly on the impurity concentration.[18]

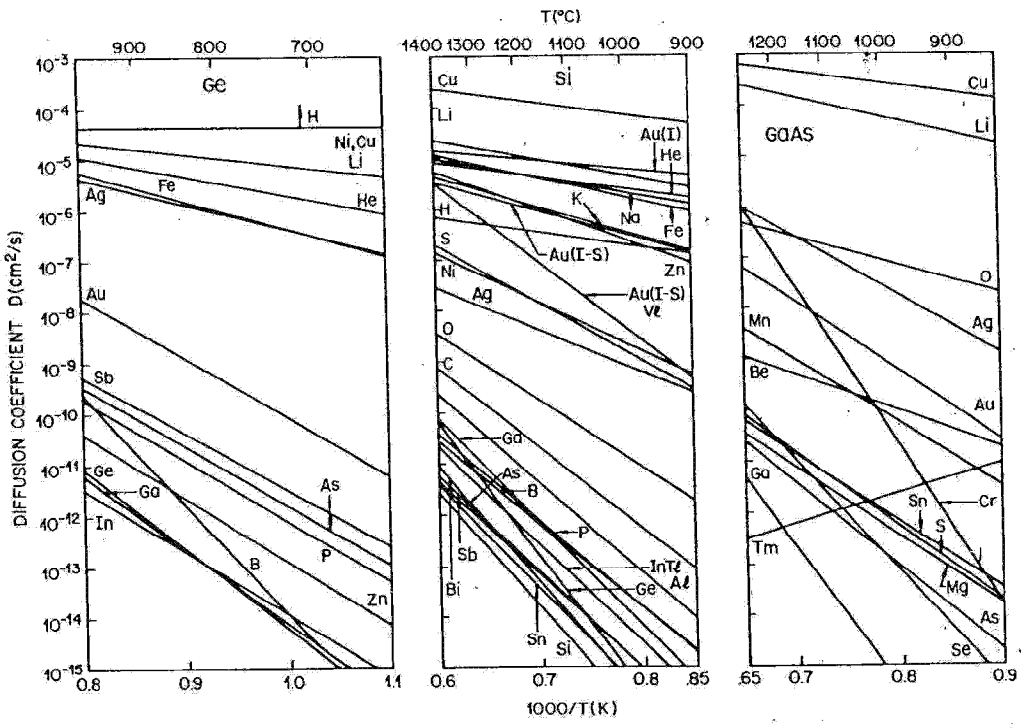

Fig. 4 Impurity diffusions coefficients as a function of temperature for Ge, Si, and GaAs. (After Burger and Donovan, Ref. 19; Kendall and DeVries, Ref. 20.)

The diffusion coefficient D depends on temperature and impurity concentration. Under low-concentration conditions, D becomes independent of impurity concentration. (In low-concentration conditions, the impurity concentration is less than the intrinsic carrier concentration at the diffusion temperature, for example, at 1100°C, $n_i \approx 10^{19}\ \text{cm}^{-3}$, as shown in Fig. 11 of Chapter 1.) In a limited temperature range and under low-concentration conditions, the diffusion coefficient can be described by

$$D(T) = D_0 \exp(-\Delta E/kT) \tag{3}$$

where D_0 is the diffusion coefficient extrapolated to infinite temperature, and ΔE is the activation energy of diffusion. Values of $D(T)$ for Ge, Si, and GaAs are plotted in Fig. 4 for various impurities.[19,20] These values are for low-impurity-concentration cases; as impurity concentration increases, $D(T)$ becomes increasingly concentration-dependent.

The impurity diffusion coefficient is related to the solid solubility of the

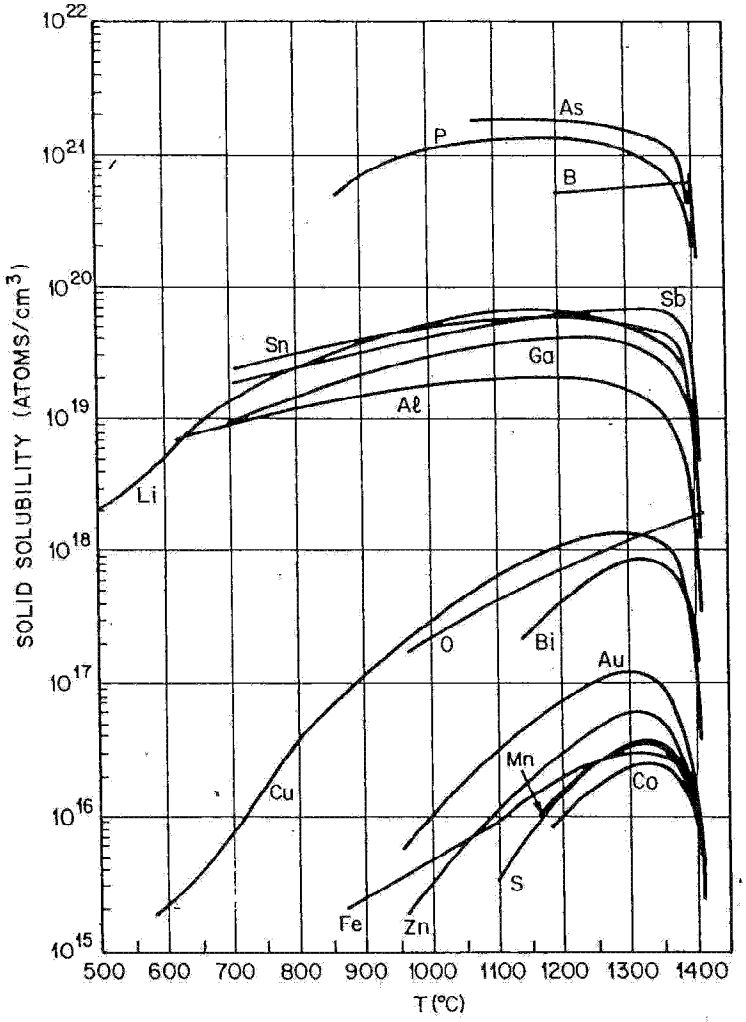

Fig. 5 Solid solubility of various elements in Si as a function of temperature. (After Trumbore, Ref. 21.)

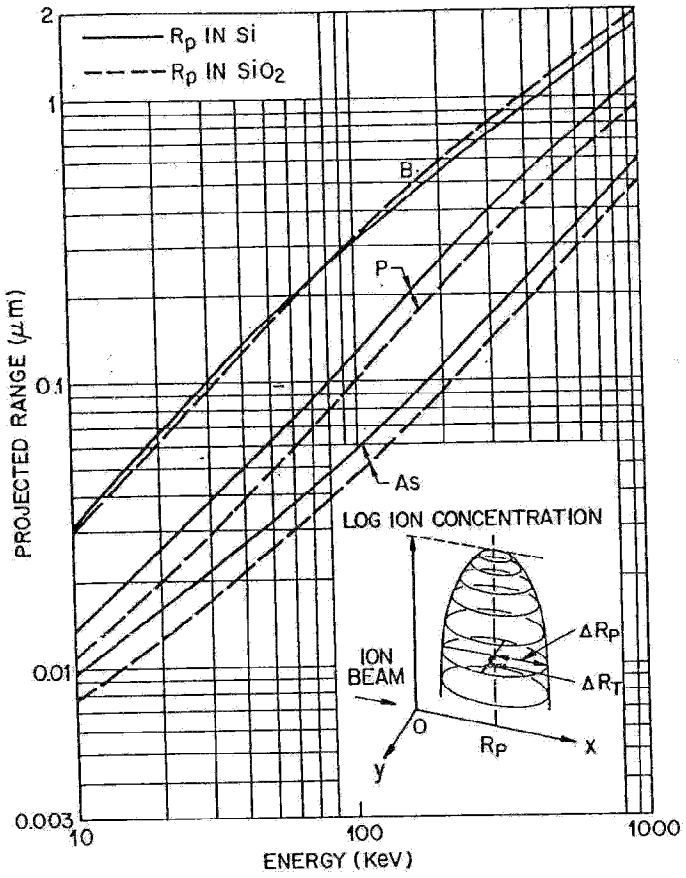

Fig. 6 Projected range of boron, phosphorus, and arsenic ions in Si and SiO_2 as a function of implantation energy. The insert shows the distribution of the implanted ions. (After Pickar, Ref. 10.)

impurity, which is the maximum concentration of an impurity that can be accommodated in a solid at any given temperature. The solid solubilities of some important impurities in Si are plotted in Fig. 5 as a function of temperature.[21] This figure shows that arsenic or phosphorus should be used as the impurity in making heavily doped *n*-type silicon, and boron should be used for heavily doped *p*-type silicon.

Ion implantation is the introduction of energetic charged atomic particles into a substrate for the purpose of changing the electrical, metallurgical, or chemical properties of the substrate. The typical ion energies considered are between 10 and 400 keV, and typical ion doses vary from 10^{11} to 10^{16} ions/cm^2. The main advantages of ion implantation include (1) precise control over total dose, depth profile, and area uniformity; (2) low-temperature processing; and (3) implanted junctions that can be self-aligned to the edge of the mask.

For an ion beam with an infinitesimally small beam diameter, the ion distribution in the substrate (Fig. 6, insert) is given by[10]

$$n(x, y) = \frac{s}{(2\pi)^{3/2}\,\Delta R_p\,\Delta R_T^2} \exp\left[-\left(\frac{x - R_p}{\sqrt{2}\Delta R_p}\right)^2\right] \exp\left[-\left(\frac{y}{\sqrt{2}\Delta R_T}\right)^2\right] \quad (4)$$

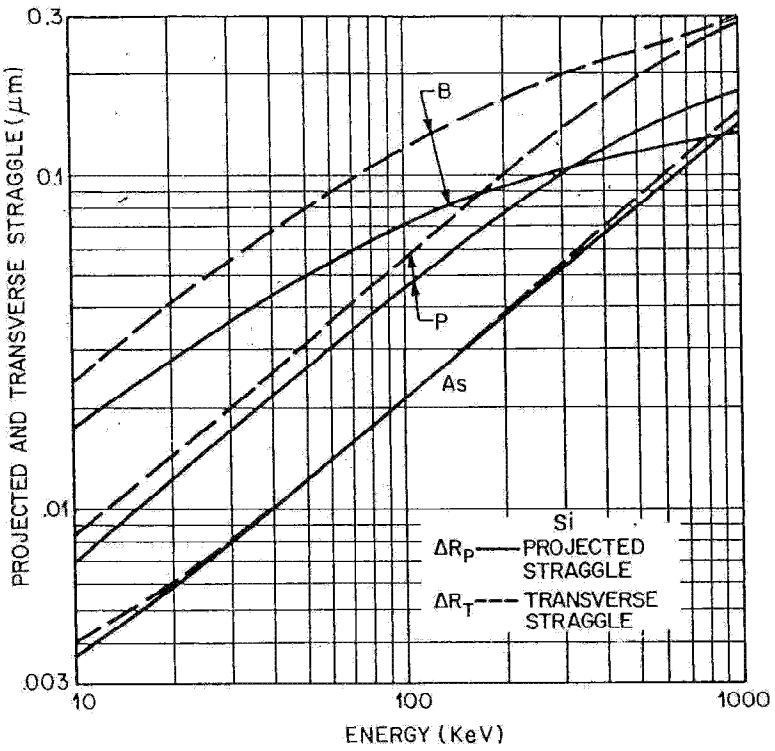

Fig. 7 Projected straggle and transverse straggle of boron, phosphorus, and arsenic in silicon as a function of implantation energy. (After Pickar, Ref. 10.)

where s is the incident rate of ions per second, R_p the projected range, ΔR_p the projected straggle, and ΔR_T the transverse straggle. Figure 6 shows values of R_p in Si and SiO_2 for boron, phosphorus, and arsenic ions as a function of implantation energy. The projected range increases approximately linearly with the energy. For boron, the projected ranges in Si and SiO_2 are quite close; however, for phosphorus and arsenic, the projected range in SiO_2 is about 20% lower than that in Si. The projected straggle and transverse straggle in Si for these ions are shown in Fig. 7. The straggles also increase with increasing energy. The ratio of ΔR_p to R_p over the energy range is about 0.2 to 0.5.

For an implantation from an infinitesimal beam that is scanned uniformly across the substrate surface, the y dependence of the doping density drops out in regions several ΔR_T from the edge of the scanned area. The doping distribution then becomes

$$n(x) = \frac{\phi}{\sqrt{2\pi}\Delta R_p} \exp\left[-\left(\frac{x - R_p}{\sqrt{2}\Delta R_p}\right)^2\right] \tag{4a}$$

which describes the Gaussian distribution in doping density of the implanted ions, where ϕ is the total number of ions per unit area, and the quantity $\phi/(\sqrt{2\pi}\Delta R_p)$ is the peak doping concentration at $x = R_p$.

Since 1974, laser processing of semiconductors has been extensively studied.[22,23] High-intensity laser radiations [such as the pulsed ruby laser

and continuous-wave (cw) argon laser] offer possibilities to remove damages associated with ion implantation and to recrystallize amorphous semiconductor layers. The potential advantages of laser processing include (1) control of the annealing depth and the impurity profile through the absorption properties of the laser light and the dwell time of the pulsed or swept beam; specifically, the laser annealing can activate the implanted impurities without impurity redistribution; (2) highly localized lateral processing on a micron scale since the laser beam can be focused down to such dimensions; and (3) regrowth of crystalline material from an amorphous layer on the crystalline substrate or formation of large-grain-size polycrystalline films deposited on insulators.

In practice, most impurity profiles can be approximated by the following two limiting cases: the abrupt junction and the linearly graded junction, as shown in Fig. 8*a* and 8*b* respectively. The abrupt approximation provides

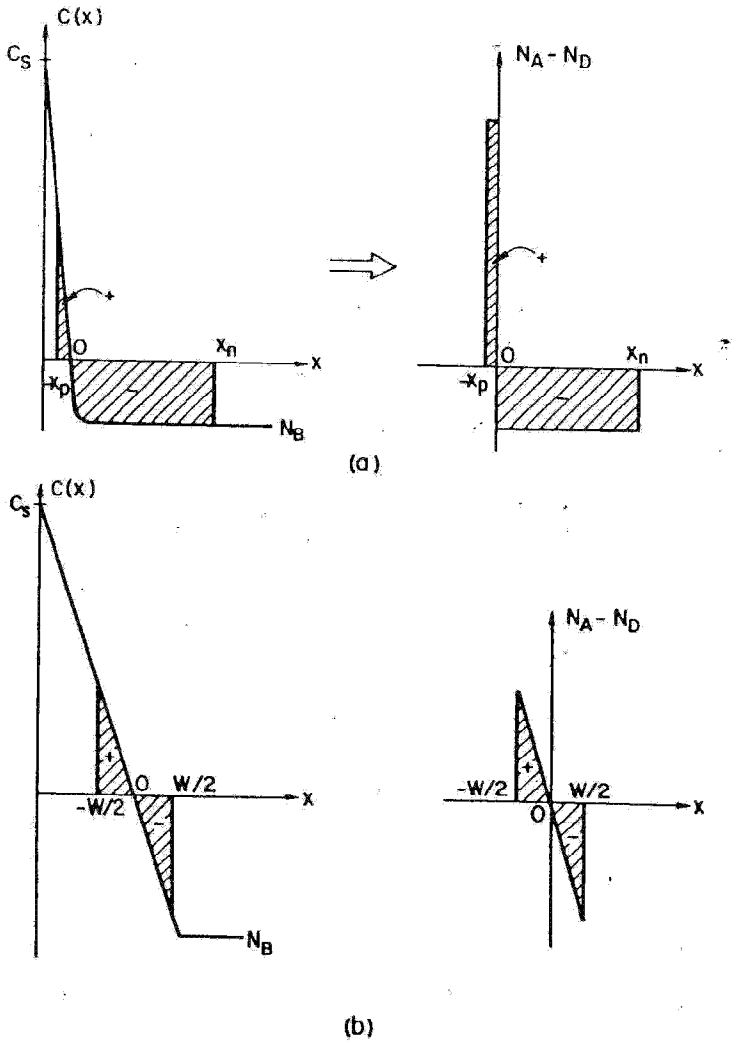

Fig. 8 Approximate doping profiles. (a) Abrupt junction. (*b*) Linearly graded junction.

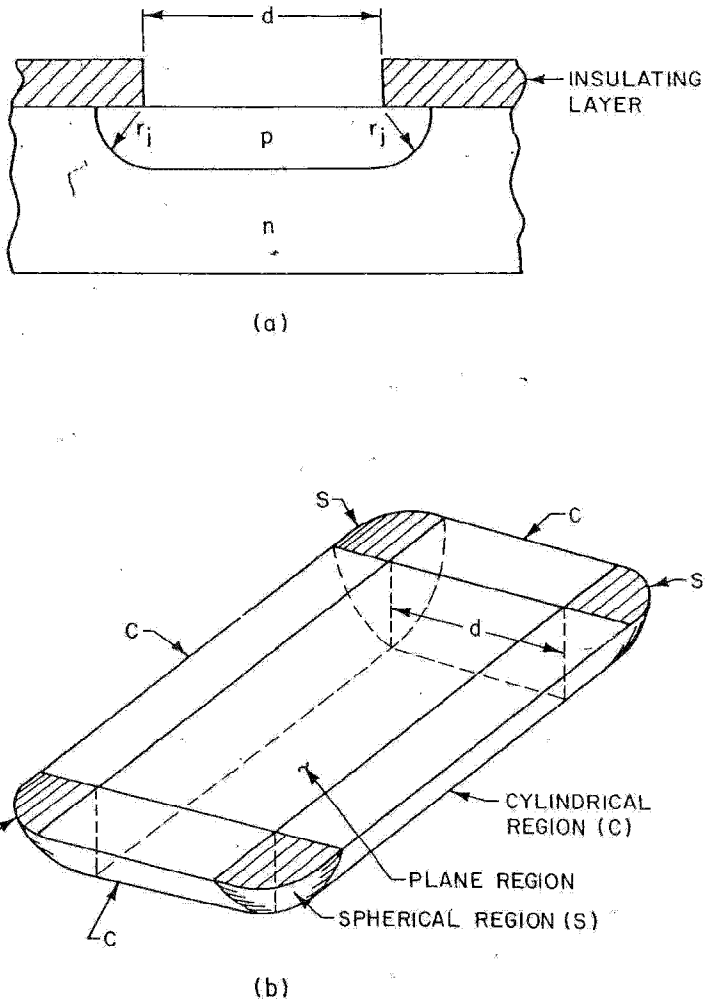

Fig. 9 (*a*) Planar diffusion process which forms junction curvature near the edges of the diffusion mask. r_j is the radius of curvature. (*b*) The formation of approximately cylindrical and spherical regions by diffusion through a rectangular mask. (After Lee and Sze, Ref. 24.)

an adequate description for alloyed junctions, shallowly diffused junctions, and ion-implanted junctions. The linearly graded approximation is reasonable for deeply diffused junctions.

Another important effect results from the planar processes. When a p-n junction is formed by diffusion into a bulk semiconductor through a window in an insulating layer, the impurities will diffuse downward and also sideways. Hence the junction consists of a plane (or flat) region with approximately cylindrical edges, as shown[24] in Fig. 9*a*. In addition, if the diffusion mask contains sharp corners, the junction near the corner will be roughly spherical in shape (Fig. 9*b*). These spherical and cylindrical regions have profound effects on the junction, especially for the avalanche multiplication process,[25] discussed in Section 2.5.

2.3 DEPLETION REGION AND DEPLETION CAPACITANCE

2.3.1 Abrupt Junction

Diffusion Potential and Depletion-Layer Width When the impurity concentration in a semiconductor changes abruptly from acceptor impurities N_A to donor impurities N_D, as shown in Fig. 10*a*, one obtains an abrupt junction. In particular, if $N_A \gg N_D$, one obtains a one-sided abrupt junction or p^+n junction.

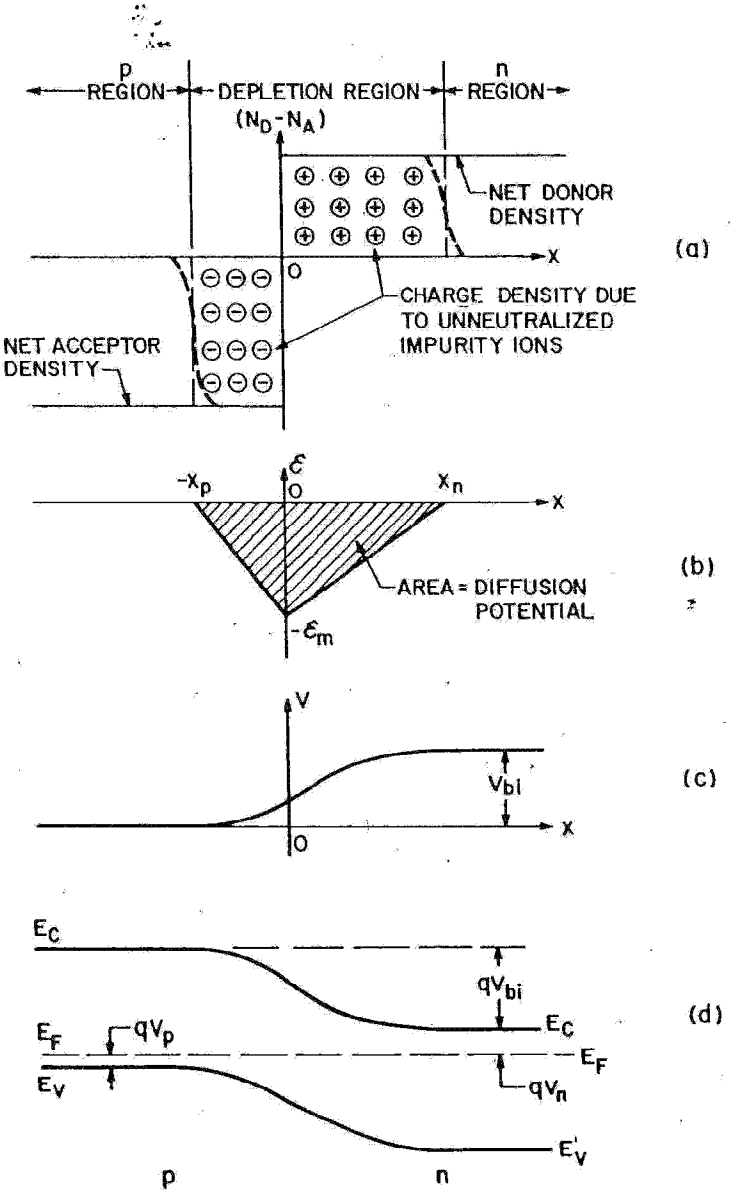

Fig. 10 Abrupt p-n junction in thermal equilibrium. (*a*) Space-charge distribution. The dashed lines indicate the majority-carrier distribution tails. (*b*) Electric field distribution. (*c*) Potential variation with distance where V_{bi} is the built-in potential. (*c*) Energy-band diagram.

We first consider the thermal equilibrium condition, that is, one with no applied voltage and no current flow. From Eqs. 33 and 93a in Chapter 1,

$$J_n = 0 = q\mu_n\left(n\mathscr{E} + \frac{kT}{q}\frac{\partial n}{\partial x}\right) = \mu_n n \frac{\partial E_F}{\partial x} \tag{5}$$

or

$$\frac{\partial E_F}{\partial x} = 0. \tag{5a}$$

Similarly,

$$J_p = 0 = \mu_p p \frac{\partial E_F}{\partial x}. \tag{6}$$

Thus the condition of zero net electron and hole currents requires that the Fermi level must be constant throughout the sample. The diffusion potential, or built-in potential V_{bi}, as shown in Fig. 10*b*, *c*, and *d*, is equal to

$$\begin{aligned} qV_{bi} &= E_g - (qV_n + qV_p) \\ &= kT \ln\left(\frac{N_C N_V}{n_i^2}\right) - \left[kT \ln\left(\frac{N_C}{n_{no}}\right) + kT \ln\left(\frac{N_V}{p_{po}}\right)\right] \\ &= kT \ln\left(\frac{n_{no}p_{po}}{n_i^2}\right) \simeq kT \ln\left(\frac{N_A N_D}{n_i^2}\right). \end{aligned} \tag{7}$$

Since at equilibrium $n_{no}p_{no} = n_{po}p_{po} = n_i^2$,

$$V_{bi} = \frac{kT}{q} \ln\left(\frac{p_{po}}{p_{no}}\right) = \frac{kT}{q} \ln\left(\frac{n_{no}}{n_{po}}\right). \tag{7a}$$

Equation 7a gives the relationship between the hole and electron densities on either side of the junction:

$$p_{no} = p_{po} \exp\left(-\frac{qV_{bi}}{kT}\right) \tag{8a}$$

$$n_{po} = n_{no} \exp\left(-\frac{qV_{bi}}{kT}\right). \tag{8b}$$

The approximate values of V_{bi} for one-sided abrupt p-n junctions in Ge, Si, and GaAs are shown in Fig. 11.

Since in thermal equilibrium the electric field in the neutral regions (far from the junction at either side) of the semiconductor must be zero, the total negative charge per unit area in the p side must be precisely equal to the total positive charge per unit area in the n side:

$$N_A x_p = N_D x_n. \tag{9}$$

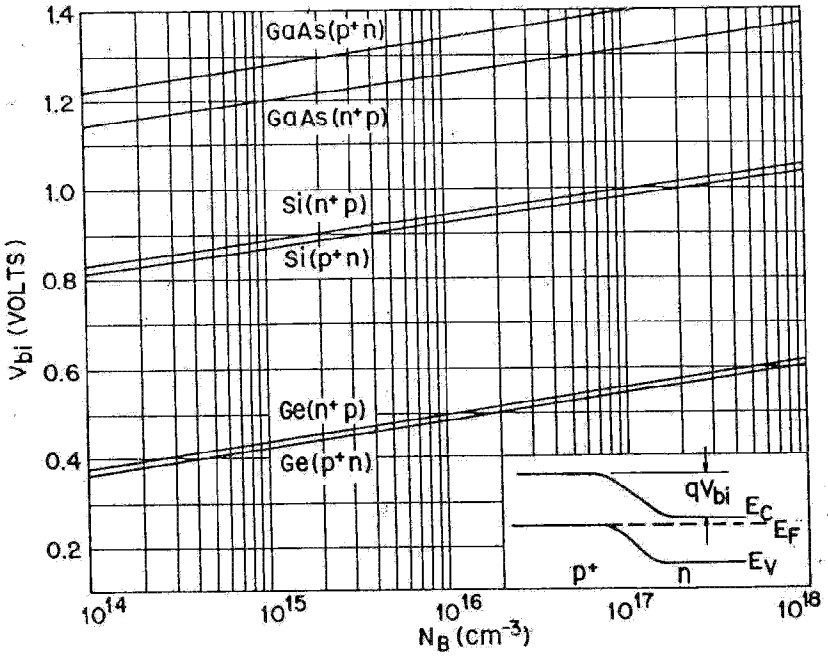

Fig. 11 Built-in potential for one-sided abrupt junctions in Ge, Si, and GaAs, where p^+ is for the heavily doped p side and n^+ is for the heavily doped n side. The background doping N_B is for the impurity concentration of the lightly doped side.

From Poisson's equation we obtain (for abrupt approximation)

$$-\frac{\partial^2 V}{\partial x^2} \equiv \frac{\partial \mathscr{E}}{\partial x} = \frac{\rho(x)}{\epsilon_s} = \frac{q}{\epsilon_s}[p(x) - n(x) + N_D^+(x) - N_A^-(x)] \tag{10}$$

or

$$-\frac{\partial^2 V}{\partial x^2} \approx \frac{q}{\epsilon_s} N_D \qquad \text{for} \quad 0 < x \leq x_n \tag{10a}$$

$$-\frac{\partial^2 V}{\partial x^2} \approx \frac{-q}{\epsilon_s} N_A \qquad \text{for} \quad -x_p \leq x < 0. \tag{10b}$$

The electric field is then obtained by integrating Eqs. 10a and 10b as shown in Fig. 10*b*:

$$\mathscr{E}(x) = -\frac{qN_A(x + x_p)}{\epsilon_s} \qquad \text{for} \quad -x_p \leq x < 0 \tag{11a}$$

and

$$\mathscr{E}(x) = -\mathscr{E}_m + \frac{qN_D x}{\epsilon_s}$$

$$= \frac{qN_D}{\epsilon_s}(x - x_n) \qquad \text{for} \quad 0 < x \leq x_n \tag{11b}$$